A DISTRIBUTION-FREE DISCRIMINATION TECHNIQUE APPLIED TO TYPICAL CINCINNATIAN (UPPER ORDOVICIAN) LIMESTONES[1, 2]

ROBERT H. OSBORNE

Department of Geological Sciences, University of Southern California, Los Angeles, California 90007

ABSTRACT

Multivariate discrimination procedures have been developed by which a given sample may be assigned to its correct population of origin on the basis of values it exhibits for the p variables observed. A discrimination method (Kendall, 1966) based on order-statistics is illustrated for possible use whenever the mathematical requisites for discriminant function analysis cannot be satisfied. The method described is distribution free, involves only counting, may be applied to variables measured on or above an ordinal scale, shows which variables are most important in the discrimination, and provides for a region of statistical indecision. The method was applied to petrographic measurements taken from 25 class 1–2 and 25 class 3 limestones collected from the Kope and Fairview Formations, exposed eastern Hamilton County, Ohio. Of eleven primary and derived variables observed, 68 percent of these limestones can be correctly assigned by the percentages of micrite, sparite, broken brachiopods, and crinoids, whereas the remaining 32 percent are included within a region of indecision. In this case, the discriminatory efficiency is not significantly increased by using more than the four variables listed above.

INTRODUCTION

Statistical discrimination methods are concerned with the assignment of a sample to its correct population of origin on the basis of the values it exhibits for each of p variables. Implicit in this statement is the assumption that the correct parent populations are known, and that there is little or no overlap of the sample space occupied by each of the k parent populations.

If the assumptions of multivariate normality and the equality of group dispersions are satisfied, the linear discriminant function for the case of two parent populations (Fisher, 1936) may be effectively used, provided there is little or no overlap of the parent populations and that each of the p variables is measured on either an interval- or ratio-scale. Stevens (1946, 1958), Krumbein and Graybill (1965), and Griffiths (1967) provide useful discussions concerning the nominal, ordinal, interval, and ratio scales of measurement. Fisher's two-population linear discriminant function can easily be extended to the case of k populations, where $k>2$ (Kendall, 1968).

Linear discriminants have been applied to many subject areas within the geological sciences. For example, Emery and Griffiths (1954), Shadle and Griffiths (1955), Grifflths (1957) Huble (1957), Wood (1961), Middleton (1962), Grifflths (1963), Potter, Shimp, and Witters (1963), Smith (1964), Grifflths (1966), and Osborne (1969) have used linear discriminants to seek solutions to a variety of petrologic problems. Chayes (1965) and Chayes and Velde (1965) applied this method specifically in igneous petrology; Reyment and Naidin (1962), Reyment (1956), and Buzas (1966) have employed linear discriminants to recognize morphological groups of foraminifera and belemnites; and Kelley (1965) used this method to distinguish differences between two forms of a marine organism each of which inhabited different environments.

Comparatively little attention has been given to nonlinear discriminant functions in the geological literature; however, quadratic discriminants do occur in the noteworthy studies by Hodges (1950), McIntyre (1962), and Harries (1965).

---

[1]Contribution Number 257, Department of Geological Sciences, University of Southern California.

[2]Manuscript received January 6, 1972.

Cooper (1963, 1965) has demonstrated that quadratic discriminants are usually robust for most unimodal symmetrical distributions, and Burnaby (1966) has described a variable transformation for use in quadratic discrimination involving two populations with unequal covariance matrices. McCammon (1969) provides a useful introduction to linear, quadratic, distribution-free, and multiple discriminants, as well as a list of relevant references.

The present study illustrates an alternate approach to the use of the two-group linear discriminant function employed by Osborne (1969). There are at least two properties of the linear discriminant which may be considered as objectionable. First, each sample is uniquely assigned to one of k populations; consequently, there is no provision for a region of statistical indecision. Secondly, there is no provision for variables measured on an ordinal scale. This second problem is extremely unfortunate, because in many geological studies it is either necessary or desirable to include at least one variable measured on an ordinal scale, perhaps in combination with other variables measured on interval- and/or ratio-scales.

Provided the data under consideration are measured on or higher than an ordinal scale, the following distribution-free discrimination method based on order-statistics may be employed where the linear or perhaps quadratic discriminant is not appropriate. This method was proposed by M. G. Kendall and is discussed in Kendall (1966) and in Kendall and Stuart (1966). This method has the following advantages: it is completely distribution free, it involves no arithmetic other than counting, it shows which variables are most important in the discrimination, and it provides for a region of statistical indecision. This method proceeds by using one variable at a time to measure its degree of discriminatory effectiveness as measured by the degree of sample overlap. It should be noted that this stepwise procedure based on order statistics is not necessarily optimal. Kendall (1966) points out that non-orthogonal cutting planes are excluded in this method and that the method fails whenever populations are non-homogeneous with respect to the variance rather than to the mean of the variables. Feldman, Klein, and Honigfeld (1969) have employed a discriminatory technique similar to the method proposed by Kendall (1966).

PROCEDURE AND RESULTS

The data on which this study is based are included in Table 1 of Osborne (1967). This data matrix (Osborne, 1967) includes twenty primary and derived petrographic variables observed on each of sixty limestones collected from the Kope and Fairview Formations in eastern Hamilton County, Ohio. Samples 1 through 25 in this data matrix were assigned to limestone classes 1–2 (Weiss and Norman, 1960; Osborne, 1967) and samples 26 through 50 to limestone class 3. Samples 51 through 60 were assigned to class 1L (Osborne, 1967), but this class will be omitted from the present study because of its relatively small sample size. Therefore, this study will be concerned with the statistical discrimination of limestone classes 1–2 and 3, using the order-statistic method suggested by Kendall (1966).

The limestones assigned to class 1–2 and class 3 consist of whole and broken fossil grains, which are either embedded in micrite or separated by sparite. Eleven petrographic variables were selected from the complete data matrix (Osborne, 1967) for use in the present study. The following variables were chosen to facilitate the comparison of the results of this study with those obtained by Osborne (1969): broken brachiopods, whole brachiopods, broken bryozoans, whole bryozoans, crinoids (disarticulated), trilobites (broken), total skeletal grains, total micrite, and total sparite. The variables pertinent to this study and their respective values are listed in Tables 1 through 4.

The data base (Osborne, 1967) for the present study requires some explanation. The petrographic identification of each of the limestone constituents represents measurement on a nominal scale. The objective of petrographic modal analysis is to estimate the volume percentage characteristic of each constituent observed in thin section. In the present case, the correct volume of each variable measured lies within about five percent on each side of the obtained value, with a 95 percent confidence (Osborne, 1967). Osborne (1967) demonstrated that there are important differences in the volume percentages of certain variables between limestone classes 1–2 and 3. Consequently, although the data were measured on a nominal scale, the relative frequencies between class 1–2 and class 3 assume a directional sense (quasi-ordinal) when trying to discriminate between these two limestone classes. The values listed for each variable in Table 1 of the earlier paper (Osborne, 1967) and in Tables 1 through 4 of this paper represent the percentage of the maximum value observed for any given variable. This data transformation was employed to assign equal weight to each variable prior to the application of the discrimination technique.

A tabulation similar to Table 1 was constructed for each of the eleven petro-

TABLE 1
Discrimination of samples based on values for micrite

Variate Value (%)	Class 1–2 Sample No.	Class 3 Sample No.
100		39
91		41
81	25	50
80		27
77		45
76		42,47
75	16	43
70		40
69		46
66		36,48
65		32
63	17	
61		30
60		28
58		38
57	21	37
55		33
54	1	34
53		35
51	20	
50	9	
48	15,4	
47		49
44		26
43	10,13	29,44
42	6	
38	24	
37		31
35	2	
34	22	
33	5	
31	12	
28	3	
26	14,18	
25	23	
24	11	
20	7	
19	8	
18	19	

graphic variables selected for this study. In such tables, the variate values observed for each variable are ranked from high to low, and are also listed under their limestone class according to sample number. Examination of such tables clearly defines regions where the two limestone classes are distinct with regard to the variate values, as well as regions where values characteristic of the two classes overlap. For example, examination of Table 1 shows that a sample with a variate value equal to or exceeding 82 percent micrite has a high probability of belonging

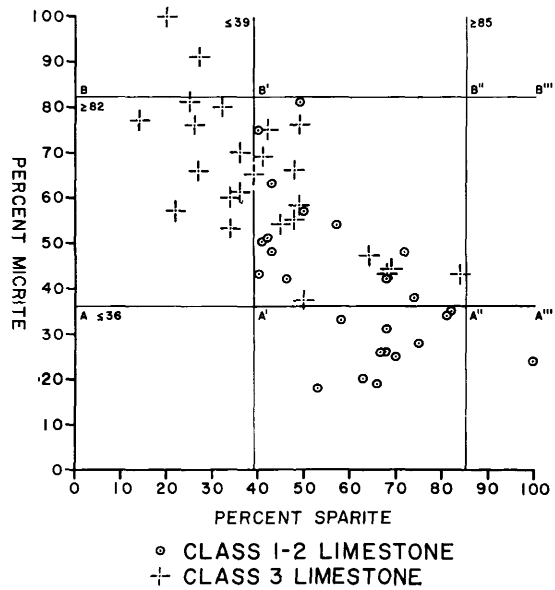

FIGURE 1. Two-dimensional diagram showing the discrimination of limestone classes 1–2 and 3 by successive use of the variables micrite and sparite.

to limestone class 3, a sample with a value equal to or less than 36 percent micrite has a high probability of belonging to class 1–2, and a sample with an intermediate value cannot be assigned to either class on the basis of micrite percentage. This assignment of limestone samples is illustrated graphically in Figure 1. Those samples with micrite values above line BB''' ($\geq$82% micrite) may be assigned to class 3, those samples with values below line AA''' ($\leq$36% micrite) may be assigned to class 1–2, and those samples between lines AA''' and BB''' cannot be assigned to either class due to the high degree of overlap in variate values typical of each class. Thus fourteen samples may be correcty assigned to their respective limestone class on the basis of values for the variable micrite. It is granted that the selection of boundary values in this example is somewhat subjective; however, boundary selection obviously becomes progressively more objective with increased sample size.

The next step is to disciminate those samples with intermediate values of micrite, i.e. those samples with values between lines AA''' and BB''' in Figure 1, by use of a second variable. The second variable selected for discrimination, namely sparite, is the one that correctly assigns the next highest number of samples. Examination of Table 2 shows that those samples with a variate value equal to or less than 39 percent for sparite have a high probability of belonging to class 3, whereas those samples with a variate value greater than 39 percent cannot be assigned. Thus, the eleven samples of limestone class 3 bounded by BB', AA',

TABLE 2
Discrimination of samples based on values for sparite

Variate Value (%)	Class 1–2 Sample No.	Class 3 Sample No.
84		29
74	24	
72	4	
69		26
68	6	44
64		49
57	1	
50	21	31
49	25	38,47
48		33,48
46	10	
45		34
43	15,17	
42	20,9	43
41		46
40	13,16	
39		32
36		30,40
34		28,35
32		27
27		36
26		42
25		50
22		37
14		45

TABLE 3
Discrimination of samples based on values for broken brachiopods

Variate Value (%)	Class 1–2 Sample No.	Class 3 Sample No.
67	9	26
66	6	
47	13	
42	10	
35		31
32		38
30	20	
26	4	
22	1	
21	16	33
20	21	
18	17	29
14	25	
13		43
12		47,49
9		46
5		48

and B'A' (fig. 1) can be effectively discriminated on the basis of sparite percentage. If a boundary were constructed at $\geq$85 percent sparite (A''B'' and the extension of this line segment), a single sample (number 11) of class 1–2 limestone could be discriminated on the basis of its sparite percentage (100 percent); however, this sample was previously assigned on the basis of its micrite percentage. The two variables, micrite and sparite, may thus be used to assign correctly 25 of the 50 limestone samples considered.

Kendall's method is an iterative procedure. Consequently, a third variable (broken brachiopods) is selected, which assigns correctly the highest number of samples with intermediate values for sparite. Table 3 and Figure 2 show that

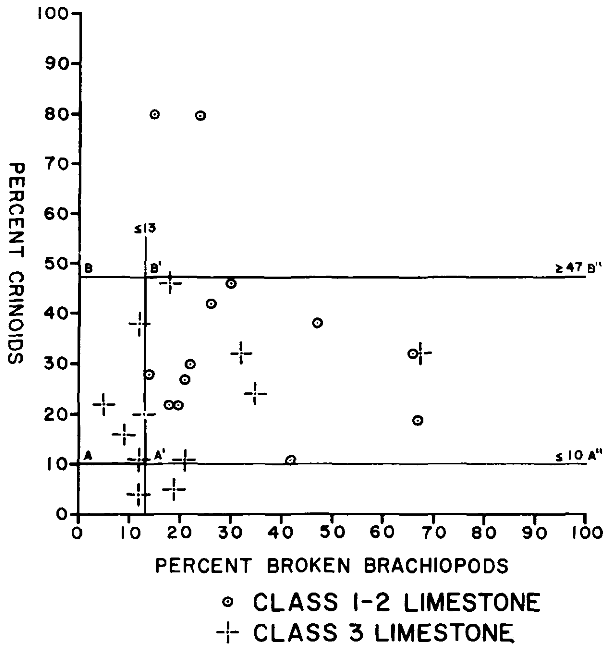

FIGURE 2. Two-dimensional diagram showing the discrimination of limestone classes 1–2 and 3 by successive use of the variables broken brachiopods and crinoids.

the variable, broken brachiopods, correctly assigns five samples to limestone class 3, whereas twenty samples remain unassigned. Similarly, Table 4 and Figure 2 show that the variable, crinoids, correctly assigns four samples to their respective limestone classes, whereas sixteen samples remain unassigned. The other seven variables included in this study provide virtually no additional discriminatory information, so these variables are not considered further. It can be seen that this procedure has efficiently assigned 34 of the fifty samples to their correct limestone class, with the remaining sixteen samples unassigned.

A discriminatory key (Table 5) may be constructed from Tables 1 through 4, which may be used to assign new samples to their correct limestone class with about a 68 percent efficiency. It is important to note that, if a sample cannot be assigned on the basis of these four variables, it simply remains unassigned rather than being misclassified.

DISCUSSION OF RESULTS

It is interesting and informative to compare the results of this study with those obtained by Osborne (1969) using a two-group, linear discriminant function.

TABLE 4
Discrimination of samples based on values for crinoids

Variate Value (%)	Class 1–2 Sample No.	Class 3 Sample No.
81	24	
80	15	
46	20	29
42	4	
38	13	49
32	6	26,38
30	1	
28	25	
27	16	
24		31
22	17,21	48
20		43
19	9	
16		46
11	10	33,47
5		34
4		44

Most of the discriminatory power using Kendall's distribution-free method is associated with micrite and sparite, whereas broken brachiopods and crinoids are of secondary importance. In the case of the linear discriminant function, however, most of the power is associated with the variables, broken brachiopods, total broken skeletal grains, total whole skeletal grains, and total skeletal grains; moderate discriminatory power is associated with micrite and sparite; and little power is associated with crinoids and trilobites. It is clear from these results that the relative importance of variables in statistical discrimination is related to the discriminatory method employed. Although the relative importance of the variables involved is an important aspect in discriminant analysis, the selection of an appropriate method may be based on considerations such as discriminatory efficiency, access to a computing facility, cost, and frequency of utilization of the method.

In the present case, the linear discriminant function is approximately ninety percent efficient (Osborne, 1969), whereas the order-statistic method is approximately 68 percent efficient with regard to this set of limestone samples. Inasmuch as misclassification will occur using both procedures, these relative efficienceies require some explanation. In the case of the two-group, linear discriminant func-

TABLE 5
Discriminatory Key

1. Micrite $\geq$ 82% assign sample to class 3
   Micrite $\leq$ 36% assign sample to class 1–2
   36% $<$ micrite $<$ 82% assignment indeterminate, refer to next variable
2. Sparite $\leq$ 39% assign sample to class 3
   Sparite $\geq$ 40% assignment indeterminate, refer to next variable
3. Broken brachiopods $\leq$ 13% assign sample to class 3
   Broken brachiopods $\geq$ 14% assignment indeterminate, refer to last variable
4. Crinoids $\geq$ 47% assign to class 1–2
   Crinoids $\leq$ 10% assign to class 3
   10% $<$ crinoids $<$ 47% assignment indeterminate

tion, approximately ten percent of any set of new samples will be assigned to the incorrect class, whereas in the order-statistic method, approximately 32 percent of the samples will simply not be assigned to either class. This region of statistical indecision associated with Kendall's order-statistic method may therefore be considered a substantive improvement. It should be noted that when a discriminant function is applied to another set of data, the observed probabilities of misclassification are generally greater than are those computed from the initial samples. Lachenbruch (1968) has shown that this increase in the probability of misclassification is directly related to the "shrinkage" of the multiple correlation coefficient in new samples.

CONCLUSIONS

In summary, the distribution-free discrimination method proposed by Kendall (1966) and illustrated in this paper may be used whenever either of the assumptions of multivariate normality or of the equality of group dispersions is violated; whenever there is considerable overlap of the parent populations; whenever there is at least one variable measured on an ordinal scale, provided all of the variables are measured at least up to an ordinal scale; and whenever statistical indecision is preferable to a relatively high level of misassignment. It is suggested that this method might profitably be employed as a scanning procedure before using more sophisticated discriminatory techniques.

REFERENCES CITED

Burnaby, T. P. 1966. Distribution-free quadratic discriminant functions in paleontology. Kansas Geol. Survey, Kansas Computer Contrib. 7: 70–77.

Buzas, M. A. 1966. The discrimination of morphological groups of *Elphidium* (foraminifer) in Long Island Sound through canonical analysis and invariant characters. J. Paleo. 40: 585–594.

Chayes, F. 1965. Classification in a ternary diagram by means of discriminant functions. Amer. Mineralogist 50: 1618–1633.

Chayes, F., and **D. Velde.** 1965. On distinguishing basaltic lavas of circumoceanic and ocean-island type by means of discriminant functions. Amer. J. Sci. 263: 206–222.

Cooper, P. W. 1963. Statistical classification with quadratic forms. Biometrika 50: 439–448.

——— 1965. Quadratic discriminant functions in pattern recognition. IEEE Trans. on Information Theory 11: 313–315.

Emery, J. R., and **J. C. Griffiths.** 1954. Differentiation of oil-bearing from barren sediments by quantitative petrographic analysis. Penn. State Univ., Min. Ind. Expt. Sta. Bull. 64: 63–68.

Feldman, S., D. F. Klein, and **G. Honigfeld.** 1969. A comparison of successive screening and discriminant function teachniques in medical taxonomy. Biometrics 25: 725–734.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann. of Eugenics 7: 179–188.

Griffiths, J. C. 1957. Petrographical investigation of the Salt Wash sediments. U. S. Atomic Energy Comm., RME 3151. 37 p.

——— 1963. Statistical approach to the study of potential oil reservoir sandstones. 3rd Ann. Conf. Computers in the Mineral Industries, Stanford Univ. Proc. 9: 637–668.

——— 1966. A genetic model for the interpretive petrology of detrital sediments. J. Geol. 74: 653–672.

——— 1967. Scientific method in analysis of sediments. McGraw-Hill Book Co., New York. 508 p.

Harries, P. D. 1965. Multivariate statistical analysis—a decision tool for mineral exploration. Short course and Symposium on Computers and Computer Applications in Mining and Exploration, College of Mines, Univ. of Ariz. 1: C1–C35.

Hodges, J. L., Jr. 1955. Discriminatory analysis, 1. Survey of discriminatory analysis. School of Aviation Med., U.S.A.F., Proj. No. 21-49-004, Rept. No. 1. 115 p.

Hulbe, C. W. H. 1957. An investigation of the size and shape of quartz grains, Pedro Beach, California. Unpub. Master's Thesis, Penn. State Univ. 69 p.

Kelley, J. C. 1965. An example of the quantitative study of echinoid morphology. Univ. Wyoming Contr. to Geology 4: 15–20.

Kendall, M. G. 1966. Discrimination and classification. p. 165–185 *in* P. K. Krishnaiah, Ed. Multivariate analysis. Academic Press, New York. 592 p.

——— 1968. A course in multivariate analysis. Griffin's Statistical Monograph No. 2, Hafner Publishing Co., New York. 185 p.

Kendall, M. G., and **A. Stuart.** 1966. The advanced theory of statistics. Vol. 3. Hafner Publishing Company, New York. 552 p.

Krumbein, W. C., and **F. A. Graybill.** 1965. An introduction to statistical models in geology. McGraw-Hill Book Co., New York. 475 p.

Lachenbruch, P. A. 1968. On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficient. Biometrics 10: 823–834.

McCammon, R. B. 1969. Discriminant analysis. p. RM–E–1 – RM–E–19 *in* P. Fenner, Ed. Models of geologic processes. Amer. Geol. Inst., Washington, D.C. 232 p.

McIntyre, D. D. 1962. A comparison of three associated environments, glacial till, fluvioglacial delta and beach sand, in terms of shapes of their quartz, garnet and hornblende grains. Penn. State Univ., Min. Ind. Expt. Sta. Publ. 78 p.

Middleton, G. V. 1962. A multivariate statistical technique applied to the study of sandstone composition. Roy. Soc. Canada Trans. 56: 119–126.

Osborne, R. H. 1967. The American Upper Ordovician Standard. VIII. R-mode factor analysis of Cincinnatian limestones. J. Sed. Petrol. 37: 649–657.

——— 1969. The American Upper Ordovician Standard. XI. Multivariate classification of typical Cincinnatian calcarenites. J. Sed. Petrol. 39: 769–776.

Potter, P. E., N. F. Shimp, and **J. Witters.** 1963. Trace elements in marine and fresh-water argillaceous sediments. Geochimica and Cosmochimica Acta 27: 669–694.

Reyment, R. A. 1966. Procedures of quantitative evaluation of variability in foraminifera. Proc. 2nd W. African Micropal. Coll. (Ibadan, 1965): 176–204.

Reyment, R. A., and **D. P. Naidin.** 1962. Biometric study of *Actinocamax versus* s.l. from the Upper Cretaceous of the Russian Platform. Stockholm Contr. Geol. 9: 147–206.

Shadle, H. W., and **J. C. Griffiths.** 1955. An attempt to establish oil-reservoir favorability criteria based on quantitative petrographic analysis. Penn. State Univ., Min. Ind. Expt. Sta. Bull. 68: 61–66.

Smith, C. M., Jr. 1964. Quantitative petrographic comparison of the Bradford third and Lewis run sands. Unpub. Ph.D. dissert., Penn. State Univ. 111 p.

Stevens, S. S. 1946. On the theory of scales of measurement. Science 103: 677–680.

——— 1958. Measurement and man. Science 127: 383–390.

Weiss, M. P., and **C. E. Norman.** 1960. The American Upper Ordovician Standard. IV. Classification of the limestones of the type Cincinnatian. J. Sed. Petrol. 30: 273–296.

Wood, G. V. 1961. Discriminating between refractory and nonrefractory quartzite by quantitative petrography. J. Sed. Petrol. 31: 530–533.